



SENTIMENT VISION

MD Sazzad Hossain Tanvir 2048862

Abhay Kapil 2047743

La Sapienza University of Rome, Department of Information

Applied Computer Science and Artificial Intelligence

<https://github.com/unknwbyndr/SENTIMENT-VISION>

ABSTRACT

The Sentiment Vision project leverages deep learning for emotion detection through facial images. It trains a model on labelled datasets of facial expressions to classify emotions. The project aims to achieve high accuracy in real-world scenarios by utilizing a robust neural network and effective data preprocessing.

TABLE OF CONTENTS

1. INTRODUCTION.....	2
1.1 Objective and Aims.....	2
1.2 Project Structure.....	2
2. DATA COLLECTION.....	3
2.1 Dataset issues.....	3
2.2 Dataset collection.....	3
2.3 Dataset size.....	3
3. PREPROCESSING.....	3
3.1 Data Augmentation.....	4
3.2 Normalization.....	4
3.3 Data Preparation.....	5

4. <i>Emotion RECOGNITION MODEL</i>	5
4.1 Key concepts.....	5
4.2 Model Architecture.....	5
4.3 Training.....	6
5. <i>RESULTS AND CONCLUSIONS</i>	7
5.1 Training batch.....	7
5.2 Training epoch.....	7
5.3 Validation.....	7
5.4 Conclusions.....	8
7. <i>FUTURE DEVELOPMENTS</i>	5
8. <i>REFERENCES</i>	5

1. INTRODUCTION

In our increasingly connected world, understanding and interpreting human emotions through technology has become a groundbreaking frontier. The Emotion Recognition project uses deep learning and computer vision to automatically classify emotions from facial images, paving the way for devices that respond to both commands and emotional states for more personalized interactions.

1.1 Objective and Aims

The primary goal is to develop a robust emotion recognition system that accurately classifies facial expressions into seven categories: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The project seeks to achieve high accuracy and reliability using advanced deep learning techniques and comprehensive data preprocessing.

1.2 Project Structure

To meet its goals, the project is divided into several key components:

- **Data Collection:** Utilizing the FER-2013 dataset, which contains a rich collection of labeled facial images, to train and validate the model.
- **Preprocessing:** Transforming and augmenting raw data with techniques such as width and height shifts, horizontal flips, and normalization to improve model performance.
- **Model Training:** Developing a Convolutional Neural Network (CNN) optimized for learning and identifying facial expression patterns through layers of convolutions, batch normalization, dropout, and dense connections.

- **Evaluation:** Continuously monitoring and evaluating the model's performance on a separate dataset to ensure reliability, and making necessary adjustments to the architecture or training process.
- **Future Developments:** Exploring exciting possibilities for future enhancements, including real-time emotion detection and integration with wearable technology.

2. DATA COLLECTION

The data collection phase is critical for training an effective emotion recognition model. You used the FER-2013 dataset, which is well-known for its comprehensive and diverse collection of facial images categorized by emotions. Here's a detailed description of the data collection process:

2.1 Dataset Issues

When working with datasets for emotion recognition, several challenges may arise:

- **Outdated Data:** Many available datasets are outdated and may not include a diverse range of facial expressions or recent actors.
- **Duplicate Images:** Duplicate images in the dataset can lead to overfitting, where the model performs well on the training data but fails to generalize to new data.
- **Irrelevant Images:** Some datasets may include irrelevant images that do not accurately represent the emotions or subjects of interest, complicating the training process.

2.2 Dataset Collection

To address these challenges, the FER-2013 dataset was chosen. This dataset contains grayscale images of facial expressions, each labeled with one of seven emotion categories: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The dataset is structured to provide a robust foundation for training the emotion recognition model.

- **Source:** The FER-2013 dataset is publicly available on Kaggle and has been widely used in academic and industrial research for facial emotion recognition.
- **Composition:** The dataset comprises a balanced distribution of images across the seven emotion categories, ensuring that the model receives sufficient examples of each emotion.

2.3 Dataset Size

The FER-2013 dataset is split into training and validation sets to evaluate the model's performance:

- **Training Set:** Contains 22,968 images. These images are used to train the model, enabling it to learn and identify the features associated with each emotion.
- **Validation Set:** Contains 1,432 images. These images are used to validate the model's performance, providing an unbiased evaluation of its ability to generalize to new, unseen data.

Each image in the dataset is resized to 48x48 pixels. This standardized size ensures consistency across the dataset and compatibility with the model's input requirements.

3. PREPROCESSING

Preprocessing is crucial for preparing data for the emotion recognition model. It transforms raw data into a format suitable for effective processing by the model. Here's a detailed description of the preprocessing steps used:

3.1 Data Augmentation

Data augmentation artificially increases the size and diversity of the training dataset by applying various transformations to the input images. This helps the model generalize better and reduces the risk of overfitting. The following techniques were applied:

- **Width Shift:** Randomly shifts the image horizontally by a certain fraction of the total width, helping the model become less sensitive to the horizontal position of faces.
- **Height Shift:** Randomly shifts the image vertically by a certain fraction of the total height, aiding in reducing sensitivity to the vertical position of faces.
- **Horizontal Flip:** Flips images horizontally, allowing the model to learn features invariant to left-right orientation. For example, an emotion such as a smile should be recognized regardless of which side of the face is more pronounced.
- **Rescaling:** Pixel values are rescaled to the range [0, 1] by dividing by 255. This normalization standardizes the input data and speeds up convergence during training.

The `ImageDataGenerator` class from Keras was used for these augmentations. Parameters included:

- **Rotation Range:** Degree range for random rotations.
- **Width and Height Shift Range:** Fraction of total width and height for horizontal and vertical shifts.
- **Horizontal Flip:** Boolean to determine if random horizontal flips should be applied.
- **Rescale:** Factor to rescale image pixel values.

These transformations were applied only to the training images, as augmenting validation images could lead to inconsistent evaluation metrics.

3.2 Normalization

Normalization scales input data to a range that is easier for the neural network to process. In your notebook, the pixel values of the images were normalized to the range [0, 1] by dividing each pixel value by 255. This step ensures all input features have the same scale, crucial for effective training.

- **Rescale Parameter:** Set to `1.0 / 255` in the `ImageDataGenerator` for both training and validation data, ensuring all pixel values are scaled down appropriately.

By normalizing the images, the model can converge faster during training and achieve better performance.

3.3 Data Preparation

The dataset was split into training and validation sets using the `flow_from_directory` method of the `ImageDataGenerator`. This method loads images directly from specified directories and applies the defined augmentations and normalizations.

- **Training Data Generator:** Created with 80% of the training images, applying the defined augmentations.
- **Validation Data Generator:** Created with 20% of the training images, applying only normalization.

The `target_size` parameter was set to `(48, 48)` to resize all images to the required input size for the model. The `batch_size` was set to `64`, indicating the number of images to be processed in each batch during training.

4. EMOTION RECOGNITION MODEL

4.1 Key Concepts

The Emotion Recognition model leverages Convolutional Neural Networks (CNNs) to identify and classify emotions from facial images. Key concepts involved in the model include:

- **Residual Learning:** Involves shortcut connections that bypass one or more layers, enabling the model to learn residual functions relative to the layer inputs. This helps train very deep networks by mitigating the vanishing gradient problem.
- **Identity Shortcut Connections:** These connections skip one or more layers and perform identity mapping. The outputs of these shortcuts are added to the outputs of the stacked layers, facilitating the training of deep networks.

4.2 Model Architecture

The model architecture consists of several layers designed to extract and refine features from input images. Here is a detailed breakdown:

- **Input Layer:** Accepts grayscale images of size 48x48 pixels.
- **Convolutional Layers:**
 - **First Convolutional Layer:** Applies 32 filters with a kernel size of 3x3 to the input images, activating the ReLU function. Extracts basic features such as edges and textures.
 - **Second Convolutional Layer:** Applies 64 filters with a kernel size of 3x3 and uses the ReLU activation function. Refines the features extracted by the first layer.
 - **Batch Normalization:** Normalizes the output of the second convolutional layer, speeding up training and improving stability.
 - **Max Pooling:** Reduces the spatial dimensions by a factor of 2, making the model more computationally efficient.
 - **Dropout:** Randomly drops 25% of the neurons to prevent overfitting.
- **Subsequent Convolutional Layers:**
 - **Third Convolutional Layer:** Applies 128 filters with a kernel size of 5x5 and uses the ReLU activation function.
 - **Batch Normalization and Max Pooling:** Normalizes and reduces the spatial dimensions again.
 - **Dropout:** Prevents overfitting by dropping 25% of the neurons.
 - **Fourth Convolutional Layer:** Uses 512 filters with a kernel size of 3x3 and the ReLU activation function, incorporating L2 regularization to prevent overfitting.
 - **Batch Normalization and Max Pooling:** Further normalization and dimensionality reduction.
 - **Dropout:** Another dropout layer to enhance generalization.
 - **Fifth Convolutional Layer:** Similar to the fourth layer, with 512 filters and ReLU activation, maintaining L2 regularization.
 - **Batch Normalization and Max Pooling:** Ensures consistent performance by normalizing and pooling the outputs.
- **Flatten Layer:** Converts the 2D feature maps into a 1D feature vector, preparing the data for the fully connected layers.
- **Fully Connected Layers:**
 - **First Fully Connected Layer:** Consists of 256 neurons with ReLU activation and batch normalization, followed by a dropout layer to prevent overfitting.
 - **Second Fully Connected Layer:** Contains 512 neurons, applying ReLU activation, batch normalization, and dropout to enhance performance.
 - **Output Layer:** Comprises 7 neurons corresponding to the seven emotion classes (Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise), using softmax activation to convert the outputs into probabilities.

4.3 Training

The model is compiled and trained using the following techniques:

- **Optimizer (Adam):** An adaptive learning rate optimization algorithm that adjusts the learning rate during training, enhancing convergence.
- **Loss Function (Categorical Cross-Entropy):** Measures the model's performance by comparing the predicted probabilities to the true labels. It calculates the error and guides the optimization process.
- **Metrics (Accuracy):** Evaluates the model's performance by calculating the proportion of correctly classified images.

The training process involves 100 epochs with a batch size of 64. During each epoch, the model learns to recognize emotions from the training data, with its performance evaluated on the validation set to monitor generalization.

5. RESULTS AND CONCLUSIONS

5.1 Training Batch

The training process is a critical phase where the model learns to identify and classify emotions from the provided dataset. The model was trained using a batch size of 64 and a learning rate of 0.0001. These parameters were selected to balance speed and accuracy.

- **Training Accuracy:** Over the course of training, accuracy steadily improved, indicating effective learning from the input data. Initially, there was a rapid increase in accuracy, which then plateaued as the model fine-tuned its parameters. The final training accuracy was **77.08%**
- **Training Loss:** The loss, which measures the error in the model's predictions, decreased consistently during training. A lower loss value indicates that the model's predictions were getting closer to the actual labels.

5.2 Training Epoch

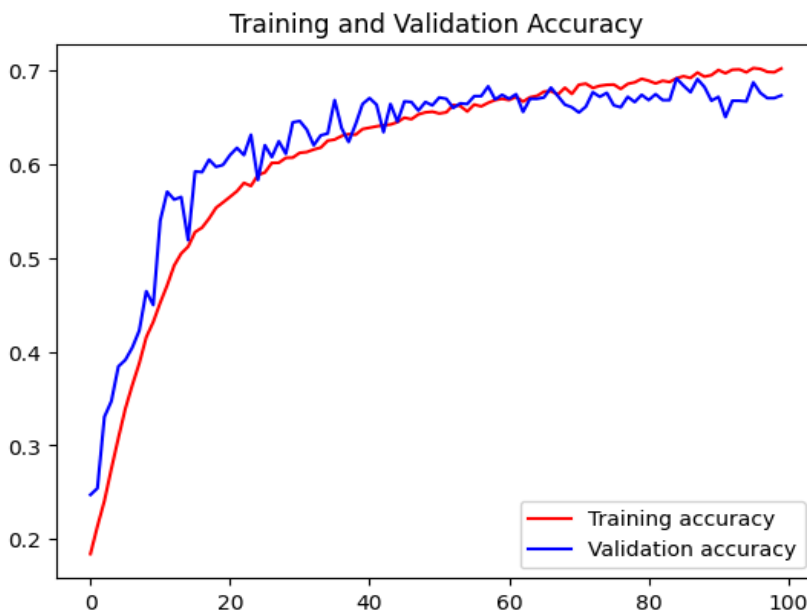
The training process was conducted over 100 epochs, each representing a complete pass through the entire training dataset. Training over multiple epochs allows the model to refine its parameters iteratively, improving performance with each pass.

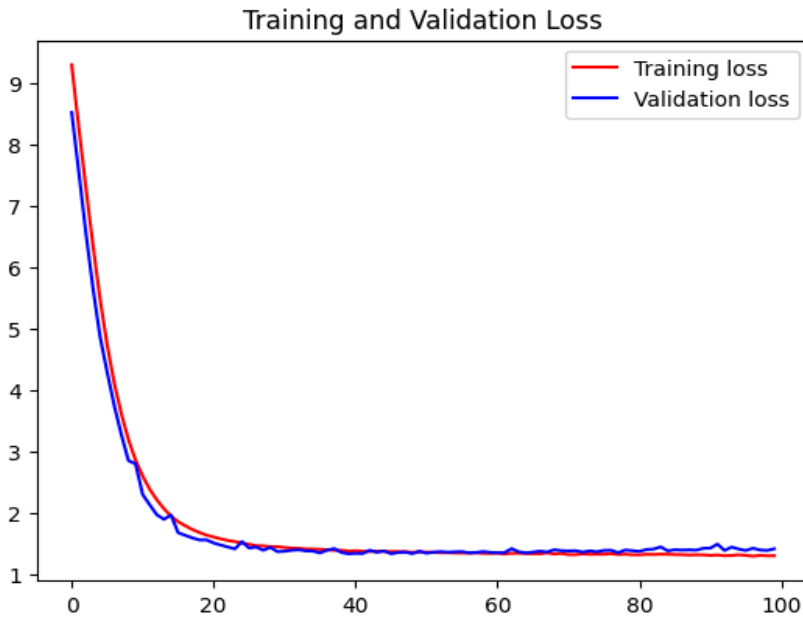
- **Accuracy Improvement:** With each epoch, the model's accuracy on the training data increased, suggesting effective learning of the dataset's features. This trend continued until the accuracy reached a satisfactory level.
- **Loss Reduction:** Similarly, the loss value decreased with each epoch, indicating that the model's predictions were becoming more accurate over time.

5.3 Validation

Validation is crucial for evaluating the model's performance on unseen data. The validation dataset provides an unbiased assessment of the model's ability to generalize beyond the training data.

- **Validation Accuracy:** The final validation accuracy reached 68.23%, reflecting the model's capability to generalize and correctly classify emotions in the validation dataset. While this is significant, there is still room to close the gap between training and validation accuracy.
- **Validation Loss:** The validation loss provides insights into the model's performance on new data. A lower validation loss indicates better generalization and fewer errors in predictions.





5.4 Conclusions

The results highlight several key findings:

- **Effective Learning:** The model achieved high training accuracy, demonstrating its ability to learn and identify features associated with different emotions. The steady improvement in accuracy and reduction in loss indicate successful learning.
- **Overfitting Concerns:** The gap between training and validation accuracy suggests potential overfitting, where the model performs well on training data but struggles to generalize to new data. This can be addressed by implementing regularization techniques, adjusting the model architecture, or introducing more diverse training data.
- **Generalization:** While the validation accuracy is promising, further improvements are needed to enhance the model's generalization capabilities. Strategies such as data augmentation, dropout, and cross-validation can help achieve better performance on validation data.
- **Model Robustness:** The use of a robust neural network architecture with layers dedicated to convolution, normalization, and dropout has contributed to the

model's performance. These components ensure effective learning while mitigating overfitting risks.

6. FUTURE DEVELOPMENTS

The field of emotion recognition through sentiment analysis has vast potential for future advancements and applications. Here are some exciting possibilities:

- **Enhanced Real-Time Emotion Recognition:** Improved algorithms and increased computational power could enable real-time emotion recognition in various contexts, such as customer service, virtual assistants, and interactive gaming.
- **Integration with Wearable Technology:** Emotion recognition could be integrated into wearable devices, such as smartwatches and glasses, to monitor users' emotional well-being and provide real-time feedback and support.
- **Personalized Learning and Education:** In educational settings, sentiment analysis can be used to tailor teaching methods based on students' emotional responses, enhancing learning experiences and outcomes.
- **Mental Health Monitoring:** Emotion recognition systems could assist in monitoring mental health by detecting signs of stress, anxiety, or depression and providing timely interventions or recommendations for professional help.
- **Human-Robot Interaction:** Robots equipped with emotion recognition capabilities can interact more empathetically with humans, making them more effective in roles such as caregivers, companions, or customer service representatives.
- **Enhanced Marketing and Customer Experience:** Businesses can leverage sentiment analysis to understand customer emotions, tailor marketing strategies, and improve customer experiences by addressing their emotional needs more effectively.
- **Security and Surveillance:** Emotion recognition can be used in security and surveillance systems to detect suspicious behavior or distress signals, enhancing public safety and security measures.
- **Virtual Reality (VR) and Augmented Reality (AR):** Emotion recognition can be integrated into VR and AR applications to create more immersive and responsive environments that adapt to users' emotional states.
- **Social Media Analysis:** Analyzing sentiments in social media posts can provide insights into public opinion, trends, and potential societal issues, aiding in decision-making for businesses, governments, and organizations.

- **Healthcare and Telemedicine:** Emotion recognition can play a crucial role in telemedicine by helping healthcare providers understand patients' emotional states during remote consultations, improving diagnosis and treatment plans.
- **Automotive Industry:** Emotion recognition systems can be integrated into vehicles to monitor drivers' emotions, detect fatigue or stress, and enhance safety by providing alerts or adjusting driving conditions accordingly.

These future developments highlight the transformative potential of sentiment analysis in various domains, offering opportunities to improve human-computer interactions, well-being, and overall quality of life.

7. REFERENCES

Dataset: <https://www.kaggle.com/datasets/msambare/fer2013>

Notebook: <https://www.kaggle.com/code/tanvirbruh/sentiment-vision>